Statistics 583, Problem Set 9 Solutions

Wellner; 5/31/2006

1. (a) Wasserman, problem 4.7.3, page 59: prove that

$$R(f(x), \widehat{f}_n(x)) \equiv E(f(x) - \widehat{f}_n(x))^2 = \text{bias}_x^2 + Var[\widehat{f}_n(x)].$$

(b) Wasserman, problem 4.7.3, page 59: Let $X_1, \ldots, X_n$ be an i.i.d. sample from a $N(\theta, 1)$ distribution with density $f_\theta(x) = \phi(x-\theta)$ where ϕ is the standard normal density function $\phi(z) = (2\pi)^{-1/2}\exp(-z^2/2)$. Consider the density estimator $\widehat{f}(x) = f_{\hat{\theta}}(x) = \phi(x - \hat{\theta})$ where $\hat{\theta} = \overline{X}_n$ is the sample mean. Find the risk of $\widehat{f}$.

Solution: (a) This is easy:

$$\begin{aligned} R(f(x), \widehat{f}_n(x)) &\equiv E(f(x) - \widehat{f}_n(x))^2 \\ &= E(f(x) - E\widehat{f}_n(x) + E\widehat{f}_n(x) - \widehat{f}_n(x))^2 \\ &= (f(x) - E\widehat{f}_n(x))^2 + 2(f(x) - E\widehat{f}_n(x))E(E\widehat{f}_n(x) - \widehat{f}_n(x)) \\ &\qquad + E(E\widehat{f}_n(x) - \widehat{f}_n(x))^2 \\ &= \text{bias}_x^2 + Var[\widehat{f}_n(x)]. \end{aligned}$$

(b) This is a good example of how asymptotic theory can give a very good indication of the behavior of risks and variances while exact calculations are quite difficult. First the simple analysis via asymptotic theory: since $\sqrt{n}(\hat{\theta}-\theta) = \sqrt{n}(\overline{X}_n - \theta) \to_d Z \sim N(0,1)$ (and, in fact $\sqrt{n}(\overline{X}_n - \theta) =_d Z$ for every n), and since $g(\theta) \equiv \phi(x-\theta)$ is differentiable with derivative $g'(\theta) = -\phi'(x-\theta) = (x-\theta)\phi(x-\theta)$ since $\phi'(z) = -z\phi(z)$, it follows from the delta method that

$$\begin{aligned} \sqrt{n}(f_{\hat{\theta}}(x) - f_\theta(x)) &= \sqrt{n}(\phi(x-\hat{\theta}) - \phi(x-\theta)) \\ &\to_d (x-\theta)\phi(x-\theta)Z \sim N(0, (x-\theta)^2\phi^2(x-\theta)). \end{aligned}$$

Thus we might expect to be able to show that

$$nR(f(x), \widehat{f}_n(x)) = nE[(f_{\hat{\theta}}(x) - f(x))^2] \to (x-\theta)^2\phi^2(x-\theta) = (x-\theta)^2 f(x)^2.$$

Now for the more difficult exact calculation. First note that

$$(\widehat{f}_n(x) - f(x))^2 = f(x)^2\left\{\frac{\widehat{f}_n(x)}{f(x)} - 1\right\}^2$$

where

$$\begin{aligned}
\frac{\widehat{f}_n(x)}{f(x)} &= \frac{\phi(x-\overline{X})}{\phi(x-\theta)} = \exp(-\frac{1}{2}(x-\overline{X})^2 + \frac{1}{2}(x-\theta)^2) \\
&= \exp\big((x-\mu)(\overline{X}-\theta) - (\overline{X}-\theta)^2/2\big) \\
&\stackrel{d}{=} \exp\left(\frac{x-\theta}{\sqrt{n}}Z - \frac{1}{2n}Z^2\right)
\end{aligned}$$

by writing

$$\begin{aligned}
(x-\overline{X})^2 &= (x-\theta+\theta-\overline{X})^2 \\
&= (x-\theta)^2 + 2(x-\theta)(\theta-\overline{X}) + (\theta-\overline{X})^2.
\end{aligned}$$

Thus

$$\begin{aligned}
&E[(\widehat{f}_n(x) - f(x))^2] \\
&= f(x)^2 E\left\{\frac{\widehat{f}_n(x)}{f(x)} - 1\right\}^2 \\
&= f(x)^2 E\left\{\exp\left(\frac{x-\theta}{\sqrt{n}}Z - \frac{1}{2n}Z^2\right) - 1\right\}^2 \\
&= f(x)^2 E\left\{\exp\left(\frac{2(x-\theta)}{\sqrt{n}}Z - \frac{1}{n}Z^2\right) - 2\exp\left(\frac{x-\theta}{\sqrt{n}}Z - \frac{1}{2n}Z^2\right) + 1\right\}
\end{aligned}$$

Now

$$\begin{aligned}
E\exp(bZ - (a/2)Z^2) &= \int_{-\infty}^{\infty} \exp(bz - (a/2)z^2)\frac{1}{\sqrt{2\pi}}\exp(-z^2/2)dz \\
&= \frac{1}{\sqrt{1+a}}\exp\left(\frac{b^2}{2(1+a)}\right). \qquad (1)
\end{aligned}$$

by a standard calculation (or by Mathematica). Thus, by using (1) first with $b = 2(x-\theta)/\sqrt{n}$, $a = 2/n$, and then with $b = (x-\theta)/\sqrt{n}$, $a = 1/n$, we find that

$$\begin{aligned}
&E[(\widehat{f}_n(x) - f(x))^2] \\
&= f^2(x)\left\{\frac{1}{\sqrt{1+2/n}}\exp\left(\frac{4(x-\theta)^2}{2n(1+2/n)}\right) - \frac{2}{\sqrt{1+1/n}}\exp\left(\frac{(x-\theta)^2}{2n(1+1/n)}\right) + 1\right\} \\
&= f^2(x)\left\{\frac{1}{\sqrt{1+2/n}}\exp\left(\frac{2(x-\theta)^2}{2(n+2)}\right) - \frac{2}{\sqrt{1+1/n}}\exp\left(\frac{(x-\theta)^2}{2(n+1)}\right) + 1\right\}.
\end{aligned}$$

Here is a plot of the exact risks (assuming the normal model is true) for $n = 20, 100, and 500$ together with the asymptotic risk (which is overlapped by the exact risk for $n = 500$):

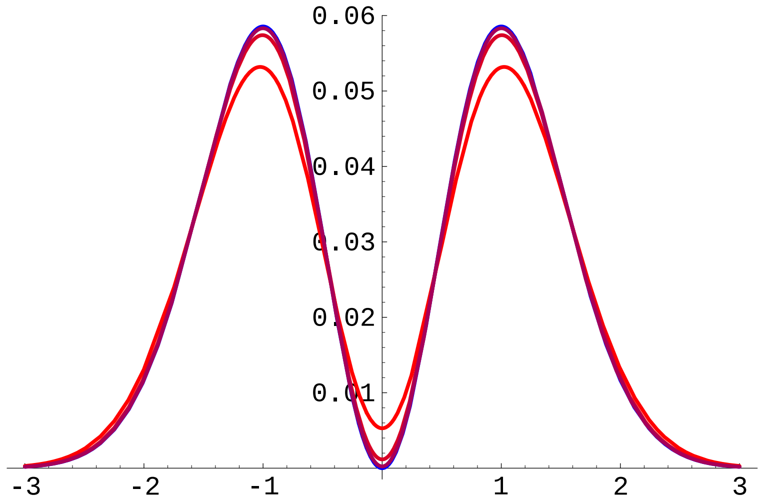

Figure 1: Exact risks of parametric density estimator for $n = 20, 100, 500$ together with asymptotic risk, as functions of x

The following plot shows the difference between the exact risk for $n = 500$ and the asymptotic risk (as a function of x):

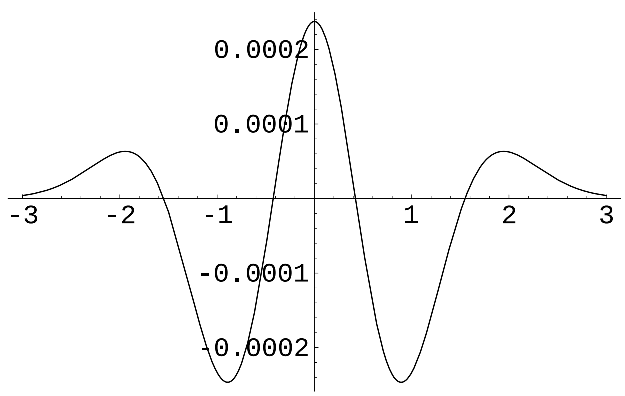

Figure 2: Difference, exact risk for $n = 500$ minus asymptotic risk of parametric density estimator

2. (a) Consider the kernel density estimator defined in (6.26), Wasserman, page 132. Show that if the density f and the kernel k satisfy the hypotheses of Wasserman's theorem 6.28, page 133, and $h = h_n$ satisfies the hypotheses of Theorem 6.27, then for fixed $x \in \mathbb{R}$,

$$\sqrt{nh_n}(\hat{f}_n(x) - E\hat{f}_n(x)) \rightarrow_d N\left(0, f(x)\int k^2(x)dx\right).$$

(b) Under what restriction on h_n does it follow (from (a) together with further analysis of the bias) that

$$\sqrt{nh_n}(\hat{f}_n(x) - f(x)) \to_d N\left(0, f(x)\int k^2(x)dx\right)?$$

(c) If $h_n = cn^{-1/5}$ and the hypotheses of (a) hold, find the limiting distribution of $\sqrt{nh_n}(\hat{f}_n(x) - f(x))$.
(d) Under the same assumptions as in (c), find the limiting distribution of $\sqrt{nh_n}(\sqrt{\hat{f}_n(x)} - \sqrt{f(x)})$.
(e) Suppose that $x, y \in \mathbb{R}$ with $x < y$. Find the joint limiting distribution of $(\sqrt{nh_n}(\hat{f}_n(x) - f(x)), \sqrt{nh_n}(\hat{f}_n(y) - f(y))$ under the assumptions in (b) and (c).

Solution: (a) First write

$$\begin{aligned}
&\sqrt{nh_n}(\hat{f}_n(x) - E\hat{f}_n(x)) \\
&= \sqrt{nh_n}\left(\frac{1}{nh_n}\sum_{i=1}^n k((x - X_i)/h_n) - \frac{1}{h_n}Ek((x - X_1)/h_n)\right) \\
&= \sum_{i=1}^n \frac{1}{\sqrt{nh_n}}(k((x - X_i)/h_n) - Ek((x - X_1)/h_n)) \\
&\equiv \sum_{i=1}^n X_{n,i}
\end{aligned}$$

where the $X_{n,i}$'s are independent and identically distributed for each n, but with a distribution depending on n. Thus we will use the Lindeberg-Feller CLT. By easy calculations, $E(X_{n,i}) = 0$, and

$$\sigma_{n,i}^2 = Var(X_{n,i}) = \frac{1}{nh_n}\int\left(k((x - y)/h_n) - \int k((x - v)/h_n)f(v)dv\right)^2 f(y)dy,$$

so

$$\begin{aligned}
\sigma_n^2 &= \sum_{i=1}^n \sigma_{n,i}^2 \\
&= \frac{1}{h_n}\int\left(k((x - y)/h_n) - \int k((x - v)/h_n)f(v)dv\right)^2 f(y)dy \\
&= \int k(z)^2 f(x - zh_n)dz - h_n\left(\int k(z)f(x - zh_n)dz\right)^2 \\
&\to f(x)\int k(z)^2 dz
\end{aligned}$$

by the dominated convergence theorem (if, for example, f is bounded). It remains only to verify the Lindeberg condition:

$$\frac{1}{\sigma_n^2}\sum_{i=1}^n E\{|X_{n,i}|^2 1\{|X_{n,i}| > \epsilon\sigma_n\} \to 0$$

as $n \to \infty$ for every $\epsilon > 0$. But

$$\begin{aligned}
X_{ni} &\equiv \frac{1}{\sqrt{nh_n}}\left\{k\left(\frac{x - X_i}{h_n}\right) - \int k\left(\frac{x-v}{h_n}\right) f(v)dv\right\} \\
&= \frac{1}{\sqrt{nh_n}} k\left(\frac{x - X_i}{h_n}\right) - \frac{h_n}{\sqrt{nh_n}}\int k(z) f(x - zh_n)dz \\
&\equiv Y_{ni} - \sqrt{\frac{h_n}{n}}(f(x) + o(1)).
\end{aligned}$$

Thus

$$\begin{aligned}
&|X_{ni}| \le |Y_{ni}| + \sqrt{\frac{h_n}{n}}(f(x) + o(1)), \\
&|X_{ni}|^2 \le 2\{Y_{ni}^2 + \frac{h_n}{n}(f(x) + o(1))^2\}, \quad \text{and} \\
&\{|X_{ni}| > \epsilon\sigma_n\} \subset \{|Y_{ni}| \ge (1/2)\epsilon\sigma_n\} \cup \{\sqrt{h_n/n}(f(x) + o(1)) > (1/2)\epsilon\sigma_n\} \\
&\qquad\qquad = \{|Y_{ni}| \ge (1/2)\epsilon\sigma_n\}
\end{aligned}$$

for n sufficiently large since the second event on the right is empty for large n. It follows that

$$\begin{aligned}
&E|X_{n,i}|^2 1\{|X_{n,i}| > \epsilon\sigma_n\} \\
&\le 2E|Y_{n,i}|^2 1\{|Y_{n,i}| > \epsilon\sigma_n/2\} + 2\frac{h_n}{n}(f(x) + o(1))^2 E1\{|Y_{ni}| \ge (1/2)\epsilon\sigma_n\} \\
&\le 2E|Y_{n,i}|^2 1\{|Y_{n,i}| > \epsilon\sigma_n/2\} + \frac{2\frac{h_n}{n}(f(x) + o(1))^2}{((1/2)\epsilon\sigma_n)^2} E|Y_{n,i}|^2 1\{|Y_{ni}| \ge (1/2)\epsilon\sigma_n\}\} \\
&= (2 + o(1))E|Y_{n,i}|^2 1\{|Y_{n,i}| > \epsilon\sigma_n/2\},
\end{aligned}$$

and hence it suffices to show that

$$\sum_{i=1}^n E|Y_{n,i}|^2 1\{|Y_{n,i}| > \epsilon\} \to 0 \qquad \text{for all } \epsilon > 0.$$

But

$$
\begin{aligned}
&\sum_{i=1}^{n} E|Y_{n,i}|^2 1\{|Y_{n,i}| > \epsilon\} \\
&= \sum_{i=1}^{n} E\frac{k((x-X_i)/h_n)^2}{nh_n} 1\{k((x-X_i)/h_n) > \epsilon\sqrt{nh_n}\} \\
&= \frac{1}{h_n}\int k((x-y)/h_n)^2 1\{k((x-y)/h_n) > \epsilon\sqrt{nh_n}\} f(y)dy \\
&= \int k^2(z) f(x-zh_n) 1\{k(z) > \epsilon\sqrt{nh_n}\} dz \\
&\to 0
\end{aligned}
$$

by the dominated convergence theorem if $\|f\|_\infty < \infty$. Thus by the Lindeberg-Feller CLT it follows that $\sum_1^n X_{n,i}/\sigma_n \to N(0,1)$. Since $\sigma_n^2 \to f(x)\int k^2(z)dz$, the conclusion follows.

(b) Note that the bias is

$$
\begin{aligned}
E\hat{f}_n(x) - f(x) &= \int k(z) f(x-zh_n)dz - f(x) = \int k(z)\{f(x-zh_n) - f(x)\}dz \\
&= \int k(z)\frac{1}{2} f''(z_x^*) z^2 h_n^2 dz
\end{aligned}
$$

where $|z_x^* - x| \le |x - zh_n - x| = |z|h_n$. Thus if $f''(x)$ is bounded and $\sqrt{nh_n}h_n^2 \to 0$, the bias term is negligible and

$$
\begin{aligned}
\sqrt{nh_n}(\hat{f}_n(x) - f(x)) &= \sqrt{nh_n}(\hat{f}_n(x) - E\hat{f}_n(x)) + \sqrt{nh_n}(E\hat{f}_n(x) - f(x)) \\
&\to_d N(0, f(x)\int k^2(z)dz) + 0 = N(0, f(x)\int k^2(z)dz).
\end{aligned}
$$

Note that $\sqrt{nh_n}h_n^2 = (nh_n^5)^{1/2} \to 0$ if and only if $nh_n^5 \to 0$; i.e. if $h_n = o(n^{-1/5})$. Thus $h_n = cn^{-1/4}$ yields negligible bias when f'' exists (and is sufficiently bounded).

(c) When $h_n = cn^{-1/5}$, then the first display in (b) yields

$$
\begin{aligned}
\sqrt{nh_n}(E\hat{f}_n(x) - f(x)) &= \sqrt{c} n^{2/5} c^2 n^{-2/5} \frac{1}{2}\int z^2 k(z) f''(z_x^*)dz \\
&\to \frac{c^{5/2}}{2} f''(x)\int z^2 k(z)dz \equiv B(c, f, k, x)
\end{aligned}
$$

Thus it follows that

$$\begin{aligned}\sqrt{nh_n}(\hat{f}_n(x) - f(x)) &= \sqrt{nh_n}(\hat{f}_n(x) - E\hat{f}_n(x)) + \sqrt{nh_n}(E\hat{f}_n(x) - f(x)) \\ &\to_d N(0, f(x)\int k^2(z)dz) + B(c, f, k) \\ &= N(B(c, f, k, x), f(x)\int k^2(z)dz).\end{aligned}$$

(d) Under the assumptions in (c), it follows from the delta-method with $g(v) = v^{1/2}$ that for any x such that $f(x) > 0$ we have

$$\begin{aligned}\sqrt{nh_n}(\sqrt{\hat{f}_n(x)} - \sqrt{f(x)}) &\to_d \frac{1}{2}f(x)^{-1/2}N\left(B(c, f, k), f(x)\int k^2(z)dz\right) \\ &= N\left(\frac{B(c, f, k, x)}{2\sqrt{f(x)}}, 4^{-1}\int k^2(z)dz\right).\end{aligned}$$

Note that the variance now depends only on the kernel and not on f.

(e) To find the joint limiting distribution of $(\sqrt{nh_n}(\hat{f}_n(x) - f(x)), \sqrt{nh_n}(\hat{f}_n(y) - f(y))$ we will use the Cramér - Wold device: let $(a, b) \in \mathbb{R}^2$. We will first show that

$$\begin{aligned}&a\sqrt{nh_n}(\hat{f}_n(x) - E\hat{f}_n(x)) + b\sqrt{nh_n}(\hat{f}_n(y) - E\hat{f}_n(y)) \\ &\quad \to_d N\left(0, a^2 f(x)\int k^2(z)dz + b^2 f(y)\int k^2(z)dz\right)\end{aligned} \tag{2}$$

for each fixed $(a, b) \in \mathbb{R}^2$. This implies that

$$\begin{pmatrix}\sqrt{nh_n}(\hat{f}_n(x) - E\hat{f}_n(x)) \\ \sqrt{nh_n}(\hat{f}_n(y) - E\hat{f}_n(y))\end{pmatrix} \to_d N_2\left(0, \begin{pmatrix} f(x) & 0 \\ 0 & f(y)\end{pmatrix}\int k^2(z)dz\right) \tag{3}$$

and will yield the desired results after analyzing the relevant bias terms. But

$$\begin{aligned}&a\sqrt{nh_n}(\hat{f}_n(x) - E\hat{f}_n(x)) + b\sqrt{nh_n}(\hat{f}_n(y) - E\hat{f}_n(y)) \\ &= \frac{1}{nh_n}\sum_{i=1}^{n}\left\{a\left\{k\left(\frac{x - X_i}{h_n}\right) - \int k\left(\frac{x - v}{h_n}\right)f(v)dv\right\}\right. \\ &\qquad \left. + b\left\{k\left(\frac{y - X_i}{h_n}\right) - \int k\left(\frac{y - v}{h_n}\right)f(v)dv\right\}\right\} \\ &\equiv \sum_{i=1}^{n}\tilde{X}_{n,i}\end{aligned}$$

where

$$\tilde{X}_{n,i} \equiv \frac{1}{nh_n}\left\{a\left\{k\left(\frac{x-X_i}{h_n}\right) - \int k\left(\frac{x-v}{h_n}\right)f(v)dv\right\}\right.$$
$$\left. + b\left\{k\left(\frac{y-X_i}{h_n}\right) - \int k\left(\frac{y-v}{h_n}\right)f(v)dv\right\}\right\}$$

We compute $E\tilde{X}_{n,i} = 0$ as in (a), but now

$$\begin{aligned}\sigma_{ni}^2 &= Var(\tilde{X}_{n,i}) \\ &= (nh_n)^{-1}\left\{a^2 nVar\{k((x-X_1)/h_n) + b^2 Var((y-X_1)/h_n)\right. \\ &\qquad + 2abCov(k((x-X_1)/h_n), k((y-X_1)/h_n))\}\end{aligned}$$

so that

$$\begin{aligned}\sigma_n^2 &= \sum_{i=1}^n \sigma_{ni}^2 \\ &= h_n^{-1}\left\{a^2\int\left(k((x-w)/h_n) - \int k((x-v)/h_n)f(v)dv\right)^2 f(w)dw\right. \\ &\quad + b^2\int\left(k((y-w)/h_n) - \int k((y-v)/h_n)f(v)dv\right)^2 f(w)dw \\ &\quad +2ab\int\left(k((x-w)/h_n) - \int k((x-v)/h_n)f(v)dv\right) \\ &\qquad \left.\left(k((y-w)/h_n) - \int k((y-v)/h_n)f(v)dv\right)f(w)dw\right\} \\ &\to a^2 f(x)\int k^2(z)dz + b^2 f(y)\int k^2(z)dz + 0\end{aligned}$$

where the argument for the first two terms is exactly as in (a), and the third term converges to 0 since

$$\begin{aligned}&h_n^{-1}\int\left(k((x-w)/h_n) - \int k((x-v)/h_n)f(v)dv\right) \\ &\qquad \left(k((y-w)/h_n) - \int k((y-v)/h_n)f(v)dv\right)f(w)dw \\ &= h_n^{-1}\int\left(k((x-w)/h_n) - h_n^2(f(x)+o(1))\right) \\ &\qquad \left(k((y-w)/h_n) - h_n^2(f(y)+o(1))\right)f(w)dw \\ &= \int k\left(z - \frac{y-x}{h_n}\right)k(z)f(y-zh_n)dz + O(h_n) \\ &\to \int k(-\infty)k(z)dzf(y) = 0\end{aligned}$$

by the dominated convergence theorem if f and k are bounded using $(y-x)/h_n \to \infty$ since $y-x>0$ and $h_n \to 0$. Since $|\tilde{X}_{n,i}| \le |a||X_{n,i}(x)|+|b||X_{n,i}(y)|$, verification of the Lindeberg condition proceeds as in (a). Thus by the Lindeberg-Feller CLT, (2) holds, and this in turn yields (3).

3. (a) Wasserman, problem 6.9.3, page 143: prove that $\widehat{J}(h)$ is an unbiased estimate of $J(h)$ for histograms and kernel density estimates.
(b) Show that (6.35) on page 136 holds.

Solution: (a) First the case of kernel density estimates: here

$$\begin{aligned} J(h) &= \int \widehat{f}_n(x)^2 dx - 2\int \widehat{f}_n(x) f(x) dx, \qquad \text{and} \\ \widehat{J}(h) &= \int \widehat{f}_n(x)^2 dx - \frac{2}{n}\sum_{i=1}^{n} \widehat{f}_{(-i)}(X_i), \end{aligned}$$

so to show that $E\{\widehat{J}(h)\} = E\{J(h)\}$, it suffices to show that

$$E\left\{\frac{1}{n}\sum_{i=1}^{n} \widehat{f}_{(-i)}(X_i)\right\} = E\left\{\int \widehat{f}_n(x) f(x) dx\right\}. \tag{4}$$

Now the right side in the last display is

$$\begin{aligned} E\left\{\int \widehat{f}_n(x) f(x) dx\right\} &= \int E\{\widehat{f}_n(x)\} f(x) dx \\ &= \int\int \frac{1}{h} k\left(\frac{x-y}{h}\right) f(y) dy f(x) dx, \end{aligned}$$

while the left side is, by conditioning on X_i,

$$\begin{aligned} E\left\{\frac{1}{n}\sum_{i=1}^{n} \widehat{f}_{(-i)}(X_i)\right\} &= \frac{1}{n}\sum_{i=1}^{n} E\left\{E(\widehat{f}_{(-i)}(X_i)|X_i)\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n} E\{\int \frac{1}{h} k\left(\frac{X_i - x}{h}\right) f(x) dx\} \\ &= \int\int \frac{1}{h} k\left(\frac{y-x}{h}\right) f(y) dy f(x) dx \\ &= \int\int \frac{1}{h} k\left(\frac{x-y}{h}\right) f(y) dy f(x) dx \end{aligned}$$

by Fubini's theorem. Comparing the last two displays yields the claim.

In the case of histogram estimators, it again suffices to show that (4) holds. But

$$\begin{aligned} E\left\{\int \widehat{f}_n(x)f(x)dx\right\} &= \int E\{\widehat{f}_n(x)\}f(x)dx = \int \sum_{j=1}^{m} \frac{p_j}{h} 1_{B_j}(x)f(x)dx \\ &= \sum_{j=1}^{m} \frac{p_j^2}{h}, \end{aligned}$$

and, on the other hand,

$$\begin{aligned} E\left\{\frac{1}{n}\sum_{i=1}^{n} \widehat{f}_{(-i)}(X_i)\right\} &= E\left\{E\{n^{-1}\sum_{i=1}^{n} \widehat{f}_{(-i)}(X_i)|X_i\}\right\} \\ &= E\{n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m} \frac{p_j}{h} 1_{B_j}(X_i)\} \\ &= \sum_{j=1}^{m} \frac{p_j}{h}\int_{B_j} f(x)dx = \sum_{j=1}^{m} \frac{p_j^2}{h}, \end{aligned}$$

and hence we conclude that (again) $E\{\widehat{J}(h)\} = E\{J(h)\}$.

(b) The claimed identity is

$$\widehat{J}(h) = \frac{1}{n^2 h}\sum_i\sum_j k^*\left(\frac{X_i - X_j}{h}\right) + \frac{2}{nh}k(0) + O(n^{-2})$$

where $k^*(x) = k^{(2)}(x) - 2k(x)$ and $k^{(2)}(z) = \int k(z-y)k(y)dy$. To check this, write

$$\widehat{f}_n^2(x) = \frac{1}{n^2h^2}\sum_i\sum_j k\left(\frac{x-X_i}{h}\right)k\left(\frac{x-X_j}{h}\right)$$

so that, with $k^{(2)}(z) = \int k(z-y)k(y)dy$,

$$\begin{aligned} \int \widehat{f}_n^2(x)dx &= \frac{1}{n^2h^2}\sum_i\sum_j \int k\left(\frac{x-X_i}{h}\right)k\left(\frac{x-X_j}{h}\right)dx \\ &= \frac{1}{n^2h^2}\sum_i\sum_j \int k\left(\frac{X_j + hy - X_i}{h}\right)k(y)hdy \\ &\qquad \text{by the change of variables } \ (x-X_j)/h = y, \quad x = X_j + hy \\ &= \frac{1}{n^2h}\sum_i\sum_j \int k\left(y - \frac{X_i - X_j}{h}\right)k(y)dy \\ &= \frac{1}{n^2h}\sum_i\sum_j k^{(2)}\left(\frac{X_i - X_j}{h}\right) \\ &\qquad \text{if we assume that } \quad k(z) = k(-z). \end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\widehat{f}_{(-i)}(X_i) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{(n-1)h}k\left(\frac{X_i-X_j}{h}\right) \\
&= \frac{1}{n(n-1)h}\sum_i\sum_j k\left(\frac{X_i-X_j}{h}\right) - \frac{1}{n-1}k(0) \\
&= \frac{1}{n^2h}\sum_i\sum_j k\left(\frac{X_i-X_j}{h}\right) - \frac{1}{n}k(0) + O_p(n^{-1})
\end{aligned}$$

Putting these pieces together yields the claimed identity, but with a remainder term which is $O_p(n^{-1})$ rather than $O_p(n^{-2})$. Thus it seems that the computational formula should really read as follows:

$$\widehat{J}(h) = \frac{1}{n^2h}\sum_i\sum_j k^{(2)}\left(\frac{X_i-X_j}{h}\right) - \frac{2}{n(n-1)h}\sum_{i=1}^{n}\sum_{j\neq i}k\left(\frac{X_i-X_j}{h}\right).$$